

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG**

Secara garis besar Tugas Akhir (TGA) merupakan perwujudan kegiatan ilmiah yang dikomunikasikan lewat bahasa tulisan. Sebagai sebuah proses transfer ilmu dan informasi dalam dunia pendidikan, semakin hari aktivitas dalam menciptakan Tugas Akhir (TGA) semakin banyak ditekuni.

Dalam menghasilkan Tugas Akhir (TGA) tersebut sering terjadi penjiplakan terhadap hasil Tugas Akhir (TGA) milik orang lain tanpa sepengetahuan pemiliknya ataupun mencantumkan sumbernya, sehingga kemiripan Tugas Akhir (TGA) pun sangat mungkin terjadi. Untuk mengatasi masalah tersebut telah banyak penelitian dilakukan terutama dalam bidang *text mining*.

Ketidakmampuan mahasiswa dalam memilih judul skripsi memang banyak sekali faktornya, salah satu diantaranya adalah karena tidak tahu atau bingung dalam menentukan tema. Ada beberapa mahasiswa yang sudah memiliki ide untuk judul skripsi mereka, akan tetapi belum dapat menentukan apakah judul tersebut akan diterima atau ditolak. Hal ini terjadi karena tidak adanya sistem yang menyediakan informasi mengenai daftar judul-judul yang telah diambil dan tidak adanya sistem yang mampu mengecek seberapa besar persentase kemiripan terhadap judul-judul yang telah ada.

Dalam menentukan diterima atau tidaknya sebuah judul skripsi yang sekarang ini dilakukan adalah dengan mengecek atau membandingkan judul tersebut dengan daftar judul-judul yang telah ada secara manual yaitu dengan melihat satu-persatu. Tentu saja hal ini akan membutuhkan waktu yang cukup lama belum lagi jika pengaju judul skripsi berjumlah banyak, tidak menutup kemungkinan waktu yang dibutuhkan juga akan semakin lama. Di samping itu, pengelolaan daftar judul-judul yang telah ada juga masih manual sehingga tidak

menutup kemungkinan data akan hilang dan sulit dalam mendistribusikan informasi tersebut kepada mahasiswa.

Salah satu metode yang dapat diterapkan dalam melakukan klasifikasi teks untuk mengetahui kemiripan suatu dokumen teks adalah metode *K-Nearest Neighbor*, yaitu sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Oleh karena itu, diperlukan suatu sistem yang mampu mendeteksi persentase kemiripan judul secara cepat dan tepat dan mampu menyediakan informasi tersebut kepada seluruh mahasiswa. Dengan sistem ini, Koordinator Tugas Akhir nantinya hanya memasukkan judul yang akan diajukan ke formulir yang telah disediakan, kemudian sistem akan mengecek secara otomatis dan menampilkan hasilnya. Hasil tersebut bisa dijadikan sebagai pertimbangan dalam menentukan apakah diterima atau ditolak judul tersebut.

Dalam penelitian ini akan dibahas tentang bagaimana cara untuk mengidentifikasi kemiripan suatu Tugas Akhir (TGA) berupa dokumen teks dengan membandingkan dokumen teks Tugas Akhir (TGA) yang sudah ada dengan tahapan yang dilakukan secara umum dalam *text mining* seperti *tokenizing*, *filtering(stoplist dan wordlist)*, kemudian akan diklasifikasikan menggunakan metode *K-Nearest Neighbor*.

Berdasarkan uraian latar belakang diatas, maka penulis tertarik mengambil tugas akhir dengan judul **“Penerapan Text Mining untuk Menentukan Klasifikasi Judul TGA menggunakan Metode KNN (K-Neight Neighbort) (Studi kasus Teknik Informatika)”**.

## 1.2 RUMUSAN MASALAH

Adapun rumusan masalah dalam penelitian ini adalah :

1. Bagaimana merancang sistem pendeteksian kemiripan judul skripsi TGA?
2. Bagaimana merancang sistem yang dapat menyediakan informasi mengenai daftar judul skripsi yang telah diambil?
3. Bagaimana penerapan algoritma *KNN* pada Sistem Pendeteksian Kemiripan Judul Skripsi?

4. Bagaimana menghitung pembobotan kata pada dokumen teks karya ilmiah menggunakan algoritma *tf/idf* (*term frequency/inverse document frequency*) dan menghitung tingkat kemiripan dokumen teks berupa karya ilmiah menggunakan metode *k-nearest neighbor*.

### 1.3 BATASAN MASALAH

Adapun batasan masalah dalam penelitian ini adalah :

1. Deteksi yang dilakukan hanya pada judul skripsi
2. Judul Skripsi yang menjadi bahan perbandingan kemiripan adalah judul yang ada di Fakultas Teknik Program Studi Teknik Informatika Universitas Malikussaleh
3. Sistem ditujukan untuk mahasiswa dan Koordinator TGA (Tugas Akhir) Fakultas Teknik Program Studi Teknik Informatika
4. Algoritma yang digunakan dalam menentukan kemiripan judul skripsi adalah algoritma *KNN*
5. Menggunakan fungsi *Rolling Hash* untuk menghitung nilai *hash* dari rangkaian *Gram* algoritma *Winnowing*.
6. Sistem dijalankan secara *online* dan dibuat dengan menggunakan bahasa pemrograman PHP dan database MySQL.

### 1.4 TUJUAN PENELITIAN

1. Untuk merancang aplikasi yang dapat mengidentifikasi kemiripan judul TGA yang satu dengan yang lainnya.
2. Untuk melakukan pengklasifikasian dokumen teks pada perangkat lunak menggunakan algoritma *k-nearest neighbor*.
3. Sistem dapat memberikan informasi mengenai daftar judul skripsi yang telah diambil

### 1.5 RELEVANSI

Pengidentifikasi kemiripan Tugas Akhir (TGA) dengan metode *k-nearest neighbor* ini diharapkan dapat membantu para peneliti khususnya di

bidang *text mining* agar dapat mengembangkannya untuk lebih baik lagi sehingga tindakan penjiplakan Tugas Akhir (TGA) dapat lebih mudah diketahui.